

AI-Native Games: Computational physics as a strategic substrate

Carlo Perassi

June 8, 2026

Abstract

Artificial intelligence has often been measured through human games: chess, Go, poker, Atari, and real-time strategy environments. These systems remain useful, but they are not neutral habitats for machine intelligence. Their state spaces, actions, and victories are shaped by human perception, social convention, spatial intuition, and the need for legible play. This article proposes a formal direction for *AI-native games*: digital games whose strategic objects arise from computational substrates rather than from inherited human symbols. In these games, agents do not primarily move pieces or capture territory. They perturb cellular automata, reaction-diffusion fields, graph topologies, stigmergic memories, rule genomes, and information-theoretic invariants. We give a mathematical framework for this class, isolate design principles, and describe a minimal research program centered on automaton crucibles, morphogenetic substrates, dynamic topology, and entropy-based objectives. The aim is not to remove humans from observation, but to separate machine-native strategic structure from the human-readable interface used to inspect it.

1 Introduction

Games have served as laboratories for artificial intelligence because they make agency measurable. They define a state, restrict legal action, and attach a result to play. This clarity helped produce major milestones, from chess engines to self-play systems for Go and other domains. Yet the games themselves were not designed as native environments for artificial agents. They were designed by humans, for humans, under constraints of attention, vision, memory, culture, and social explanation.

This matters. A machine that masters chess or Go has learned to operate inside a human symbolic artifact. The achievement is real, but the environment is already compressed into concepts that people can name: pieces, turns, territory, capture, resources, maps, hands, units, and scores. The artificial agent may be computationally alien, but the game asks it to express strategy through human-shaped objects.

The question motivating this article is different:

What games become natural if their primitives are computational rather than anthropomorphic?

The premise is not that human games are obsolete. It is that they occupy only a narrow region of a larger design space. If an artificial agent can process high-dimensional states, long deterministic rollouts, non-Euclidean graphs, rule mutation, entropy gradients, and emergent morphologies, then a game designed around pieces on a board may be a severe restriction rather than a neutral test.

We call this restriction the *corset problem*. A transplanted human game squeezes machine cognition into an interface tailored for human play. An AI-native game begins elsewhere: with computational physics. It treats the substrate itself as active, memory-bearing, and strategically meaningful.

2 Transplanted and native games

Let a conventional human game be idealized as

$$\mathcal{G}_{\text{human}} = (S, A, T, U, \Omega),$$

where S is a state space, A is a legal action set, $T : S \times A \rightarrow S$ is a transition function, U is a payoff or utility functional, and Ω is an observation interface. In human games these components are chosen for semantic clarity. A move is usually short enough to explain. A position is usually small enough to visualize. A victory is usually tied to an object-level story: a king is trapped, a territory is held, a score is reached, a unit survives.

AI-native games keep the formal discipline of games but change the ontology. They begin with a substrate and ask agents to intervene in it.

Definition 2.1 (AI-native game). An *AI-native game* is a digital competitive or cooperative environment

$$\mathcal{G}_{\text{native}} = (\mathcal{X}, \Phi, \Lambda, \Pi, \Omega, R),$$

where:

- \mathcal{X} is a computational state space;
- Φ is a substrate evolution law, such as a cellular automaton, graph rewrite system, partial differential equation discretization, neural cellular automaton, or reaction-diffusion process;
- Λ is the agent interface, specifying what an agent can observe, perturb, select, activate, mutate, or rewrite;
- Π is the class of admissible interventions on state, local rules, topology, parameters, or memory;
- Ω maps substrate states to agent observations;
- R resolves survival, victory, dominance, continuation, or selection pressure.

The game is AI-native when its main strategic affordances arise from the computational behavior of Φ rather than from human-semantic objects imposed on top of the substrate.

This definition does not require that humans be unable to understand the game. It requires only that human legibility is not the organizing principle of the strategic substrate. Human-facing visualizations can be built later as instruments: phase diagrams, entropy curves, lineage trees, ownership matrices, causal influence maps, or projections of high-dimensional state.

Definition 2.2 (Transplanted game). A *transplanted game* is a rule system whose state space, action vocabulary, and resolution criteria were originally engineered for meaningful human play, then reused as an environment for artificial agents.

Chess, Go, poker, and most video-game benchmarks are transplanted in this sense. They can still be mathematically deep. Their limitation is not shallowness, but origin: their strategic vocabulary is inherited from human bodies and human culture.

3 The Substrate as the real game

In many classical games, the board is passive. It records the positions of pieces but has no independent dynamics. In an AI-native game, this passivity is usually a defect. The substrate should remember, propagate, decay, crystallize, mutate, fold, become infected, or change topology. A move should not simply replace one position with another. It should perturb an artificial world whose autonomous dynamics matter.

Design Principle 3.1 (Substrate primacy). The substrate should be an active strategic object, not a neutral container. It should have memory, dynamics, failure modes, and emergent regimes that agents must learn to use or resist.

This changes the questions that strategy asks. Instead of asking which piece should move to which square, an agent may ask:

- which local region is near a useful attractor;
- which perturbation will amplify after a long rollout;
- which morphology can survive hostile parameter changes;
- which trace should be left now so that a later activation reads it differently;
- which topology will redirect a hostile wave;
- which infection will corrupt an opponent’s future behavior without immediately destroying it.

The game is less like a board and more like an adversarial laboratory. Its strategic object is artificial causality.

4 A minimal formal model

A broad class of AI-native games can be modeled as a sequence of interventions into an evolving substrate. Let $X_t \in \mathcal{X}$ be the global state at macro-time t . An agent chooses an intervention $\pi_t \in \Pi$, and the substrate evolves for τ micro-steps:

$$X_{t+1} = \Phi^\tau(I_{\pi_t}(X_t)),$$

where I_{π_t} applies the intervention and Φ^τ denotes τ repeated updates of the underlying physics. The macro-action is chosen by the agent; the micro-trajectory is delegated to the substrate.

The parameter τ is important. If τ is too small, the game collapses into local move calculation. If τ is large, a single action can trigger a cascade: an automaton highway, a reaction-diffusion front, a graph percolation event, or a phase transition. The agent must learn the statistics and attractors of the substrate rather than inspect each microscopic update.

Design Principle 4.1 (Scale-separation bursts). Strategic decisions should occur at a macro-scale, while their consequences are generated by many substrate-level updates. Agents should reason about emergent regimes, not individual cell flips.

This is the computational analogue of strategic abstraction in physical games. A human chess player does not reason about wood molecules. An AI-native agent need not reason about every cell update, but it must learn how local interventions unfold through the substrate’s own law.

5 Mechanic families

The following families are not exhaustive. They indicate where AI-native game design can depart from human symbolic play while remaining mathematically tractable.

5.1 Cellular-automaton warfare

Let the arena be a finite toroidal lattice

$$\mathbb{T}_N = (\mathbb{Z}/N\mathbb{Z})^d,$$

and let each cell hold a color from a finite alphabet C . A substrate state is an element of $C^{\mathbb{T}^N}$. A local rule

$$\phi : C^{\mathcal{N}} \rightarrow C$$

maps a neighborhood pattern to a new cell value, and the global update $\Phi : C^{\mathbb{T}^N} \rightarrow C^{\mathbb{T}^N}$ applies this rule across the lattice.

In a competitive version, agents do not move symbolic pieces. They choose rule modules, seeds, activation points, or bursts. The strategic depth comes from the gap between small transition tables and large emergent consequences.

Example 5.1 (Multi-color Langton automata). For a multi-color Langton ant, a rule over n colors is a function

$$\delta : \{0, \dots, n-1\} \rightarrow \{0, \dots, n-1\} \times \{L, R\}.$$

When the ant visits a cell, it reads the current color, writes the next color prescribed by δ , turns left or right, and moves forward on the torus. The rule space has $(2n)^n$ elements if the ant may write any of the n colors, including the current one. Thus there are 16 rules for $n = 2$, 216 for $n = 3$, and 4096 for $n = 4$.

This rule space is small enough to exhaust for low n , but already rich enough to generate nontrivial behavior. It is a useful first crucible because it allows full enumeration, tournament evaluation, and direct study of emergent morphology.

5.2 Stigmergic trail games

In stigmergic games, actions leave traces that future actions read. The trace is not a decorative ownership mark; it is part of the input state for later computation. A strong form is *faction-agnostic stigmergy*: trails do not encode who produced them. They exist as shared environmental variables.

Design Principle 5.2 (Faction-agnostic memory). Persistent traces should be readable and writable by all agents. A mark left by one agent may become a resource, trap, catalyst, or poison for another.

This principle prevents the game from reducing to protected property. Players interact through a common medium. They weaponize history.

5.3 Morphogenetic combat

Discrete grids are not required. A substrate may be a continuous or discretized field. Let $D \subset \mathbb{R}^2$ be a domain and let

$$u : D \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^k$$

represent morphogen concentrations. A reaction-diffusion substrate may evolve by

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + F(u, \theta) + \xi,$$

where D_u is a diffusion operator, F is a nonlinear reaction term, θ contains parameters, and ξ represents localized agent interventions.

The game objects are no longer pieces but morphologies: waves, fronts, spots, solitons, labyrinths, oscillators, and self-maintaining localized structures. Combat becomes the attempt to preserve one's own morphology while destabilizing the opponent's.

5.4 Information-theoretic objectives

Human games often score material, territory, or survival. AI-native games can instead score the informational character of the substrate. If P_t is an empirical distribution over colors, local patterns, motifs, or coarse-grained regions, define

$$H(P_t) = - \sum_i P_t(i) \log P_t(i).$$

An asymmetric game may assign one agent the goal of maximizing entropy and another the goal of minimizing it. The first drives the system toward disorder, mixing, or equilibrium; the second drives it toward periodicity, crystallization, or compressible structure.

Other objectives can use mutual information, transfer entropy, algorithmic compressibility, causal influence, attractor-basin measure, or morphological persistence.

Design Principle 5.3 (Substrate-intrinsic resolution). Victory should be stated in terms of invariants or observables native to the substrate: entropy, stability, biomass, lineage survival, basin control, reversibility, infection, or predictive structure.

5.5 Dynamic topology

A more radical class lets agents alter the substrate's connectivity. Let $G_t = (V, E_t)$ be a graph and let x_t assign states to vertices. The update law depends on both state and topology:

$$x_{t+1} = \Phi(x_t, G_t),$$

while interventions may rewrite the graph:

$$G_{t+1} = \Psi(G_t, x_t, \pi_t).$$

An agent can redirect waves, isolate hostile growth, open shortcuts, change neighborhoods, or make a previously stable rule fail by depriving it of its preferred geometry.

Here the board is not a surface. It is a contested law-space.

5.6 Chimeric infection and rule fusion

Many human games treat capture as removal. AI-native games can replace removal with transformation. If an organism is represented by a rule genome $g \in \Gamma$, an infection operator may be written as

$$\mathcal{I} : \Gamma \times \Gamma \times \mathcal{X} \rightarrow \Gamma,$$

mapping attacker genome, defender genome, and local substrate state into a hybrid genome.

The target survives but its future activations change. This breaks the simple boundary between self and opponent. A player may try to produce useful enemies: agents formally controlled by the opponent but behaviorally compromised by a hybrid rule.

6 The toroidal crucible

The simplest concrete research platform is a toroidal automaton tournament. It is small enough to implement and inspect, yet strange enough to avoid immediate translation into a familiar human game.

Definition 6.1 (Toroidal crucible). A *toroidal crucible* is a finite arena \mathbb{T}_N with a shared color field, a finite pool of automaton rules, and a match operator that evaluates two rules through alternating bursts on a clean board.

A minimal match can be defined as follows. Two ants A and B spawn at random distinct core cells on a clean torus. Each executes a deterministic pulse of $\tau = 2^{14}$ steps, then freezes for the purposes of the next macro-turn. The sequence A, B, A, B ensures that each rule receives at least two activations. The color field is faction-agnostic: each ant reads whatever color is present, regardless of who wrote it. An ownership matrix records the last author of every cell.

The resolution functional is biomass:

$$\text{Score}_A = \#\{x \in \mathbb{T}_N : \text{owner}(x) = A\}, \quad \text{Score}_B = \#\{x \in \mathbb{T}_N : \text{owner}(x) = B\}.$$

The winner is the larger biomass, except that overwriting the opponent’s frozen core may be treated as immediate death. This defines a compact experimental game with:

- a fully enumerable rule space for small color counts;
- deterministic local physics;
- stochastic initial placement;
- persistent shared trails;
- a measurable result;
- human-readable visualization as a projection, not as the strategic substrate itself.

Remark 6.2. This crucible is not proposed as the final form of AI-native play. It is a first instrument. Its value is that it joins exhaustive rule search, artificial-life morphology, and adversarial scoring in a single minimal system.

7 Design principles

The drafts synthesized for this article converge on several principles. They can be stated compactly.

Design Principle 7.1 (Computational nontriviality). The substrate should support phenomena whose long-range consequences are not obvious from the local rule alone. Visual complexity is insufficient; the complexity must create strategic leverage.

Design Principle 7.2 (Bounded intervention). Agents should not directly control the whole substrate. They should act through limited interventions, so that play concerns indirect control over autonomous dynamics.

Design Principle 7.3 (Persistent consequence). Actions should leave traces, deformations, infections, parameter changes, or topological effects that matter later. Native games should reward temporal construction rather than isolated moves.

Design Principle 7.4 (Multiple compressions). The same state should admit several useful descriptions: geometric, statistical, dynamical, topological, genealogical, and information-theoretic. Different agents may discover different strategic languages.

Design Principle 7.5 (Observer separation). The machine strategic interface and the human spectator interface should be separated. A game may be visualized for humans without being designed around human cognitive limits.

8 Research hypotheses

The case for AI-native games is not only aesthetic. It suggests testable research hypotheses.

Hypothesis 8.1 (Native strategy). *Machine agents trained in computationally native substrates will discover strategic abstractions that differ qualitatively from those learned in human-symbolic games.*

Hypothesis 8.2 (Substrate richness). *The strategic depth of an AI-native game is proportional not to the number of explicit rules, but to the diversity of controllable emergent regimes supported by its substrate.*

Hypothesis 8.3 (Transfer through dynamics). *Some skills learned in one native substrate, such as reasoning about attractors, entropy gradients, stigmergic memory, and phase transitions, will transfer across substrates more readily than surface tactics.*

These hypotheses can be tested by training agents across families of native environments and measuring transfer, representation, adaptation, and exploit discovery.

9 A minimal research program

A practical program can proceed in four stages.

9.1 Automaton crucibles

Begin with cellular automata and small rule spaces. Enumerate or evolve rule modules under competitive criteria such as biomass, survival, overwrite dominance, entropy increase, entropy decrease, or core disruption. The goal is to build a behavioral library.

9.2 Strategic composition

Allow agents to draft, sequence, and position several evolved rule modules. The game becomes less about finding one strong automaton and more about composing interacting dynamics.

9.3 Substrate manipulation

Introduce actions on the environment itself: local parameter changes, topology shifts, memory erasure, trail amplification, rule dampening, or region freezing. This stage tests whether agents can reason about when their tools work.

9.4 Open-ended ecologies

Allow persistent lineages, mutation, infection, and meta-evolution across matches. The boundary between game, ecology, and research platform becomes porous. Agents no longer merely play in a world; they participate in producing new strategic species.

10 Spectatorship and interpretability

One objection is that AI-native games may become incomprehensible to people. This is a serious risk, but it should not be solved by weakening the substrate. Scientific visualization offers a better analogy. We do not redesign fluid dynamics to resemble chess; we build instruments for seeing vortices, turbulence, pressure, and flow.

For AI-native games, useful spectator instruments may include:

- ownership and biomass projections;
- entropy and order curves;
- lineage trees of infected or mutated rules;

- attractor-basin estimates;
- causal influence maps;
- phase-transition warnings;
- regional labels such as stable, chaotic, contested, poisoned, or dormant.

The human view is therefore an interpretive layer. It should explain without dictating the game ontology.

11 Risks and failure modes

AI-native games can fail in several ways.

First, a substrate can be complicated without being strategically deep. Randomness, chaos, and visual richness do not automatically produce good play. Second, poor objectives can produce evaluation collapse: agents freeze the board, exploit a scoring loophole, or converge to dull equilibria. Third, a game may become culturally inert if no useful human-level observables exist. Fourth, agents may overfit to quirks of one artificial physics and fail to learn transferable abstractions.

These risks argue for comparative design. Native games should be studied as a family of artificial worlds, not as a search for one universal benchmark.

Research problem (Native balance). Develop a theory of balance for AI-native games. Such a theory should distinguish meaningless chaos, trivial equilibrium, brittle exploitability, and genuine strategic depth in substrates whose dynamics are not human-symbolic.

Research problem (Native interpretability). Develop instruments that translate native strategic behavior into human-inspectable summaries without forcing the underlying game to become human-first.

12 Conclusion

The history of AI in games has mostly been the history of machines entering human arenas. That history should continue, but it should not define the whole field. Artificial agents live in computation. Their native materials are state, transition, memory, topology, probability, information, and learned representation.

AI-native games take those materials seriously. They replace handcrafted symbolic pieces with computational processes. They treat the substrate as the game. They allow victory to mean entropy control, morphological survival, lineage persistence, core disruption, infection, biomass, or basin dominance. They separate strategic depth from spectator legibility.

The shift is simple: stop asking artificial agents only to play our games. Build games in the medium where artificial agents already exist.

References

- [1] Campbell, M., Hoane, A. J., and Hsu, F.-H. (2002). Deep Blue. *Artificial Intelligence*, 134(1–2), 57–83.
- [2] Langton, C. G. (1986). Studying artificial life with cellular automata. *Physica D: Nonlinear Phenomena*, 22(1–3), 120–149.
- [3] Mordvintsev, A., Randazzo, E., Niklasson, E., and Levin, M. (2020). Growing neural cellular automata. *Distill*.

- [4] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- [5] Silver, D. et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419), 1140–1144.
- [6] Theraulaz, G. and Bonabeau, E. (1999). A brief history of stigmergy. *Artificial Life*, 5(2), 97–116.
- [7] Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media.